SCENE CLASSIFICATION APPARATUS OF VIDEO

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a scene classification apparatus for analyzing uncompressed or compressed video and classifying them into various types of scenes, and particularly relates to a scene classification apparatus for enabling efficient searching and classifying of, and access to the video.

Description of the Related Art

As a prior technique relating to the scene classification of video, for example, a system for inputting video of television broadcasting and classifying them by comparatively large units such as news, sports and commercial is suggested. After not only characteristics of the video but also characteristics of audio data accompanied by the video are taken into consideration, a scene classification method is suggested. As to detection of a highlight scene as summary information, a technique for extracting a highlight scene of a sports video in a compressed domain of the compressed video using the audio characteristics accompanied by the data is suggested.

In most of the prior techniques, mainly the video and the audio data accompanied by them are analyzed in an uncompressed data domain, and thus compressed video should be once subject to a decoding process. There, therefore, arises a problem that high processing costs and much processing time are necessary.

Since a unit of the scene classification is mainly comparatively large, a scene classification technique according to a more detailed unit is not established. The classification by a detailed unit is important and effective, for example, in viewing a specified scene in video and classifying in a video database.

In a highlight scene extracting method using the conventional audio characteristics, since a peak level of audio data is evaluated, if a plurality of peaks exist in a certain short time interval, an overlapped interval may be extracted as the highlight scene. Since a commercial in television broadcasting occasionally has a comparatively high audio level, the commercial may be misdecided as the highlight scene.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a scene classification apparatus for classifying uncompressed or compressed video into various types of scenes at low cost and with high accuracy using characteristics of a video and audio characteristics accompanied by the video.

In order to achieve the above object, the present invention is such that following measures are taken in a scene classification apparatus for segmenting video into shots and classifying each scene composed of one or more continuous shots based on a content of the scene.

(1) The apparatus is provided with: a detector for detecting shot density DS of the video; a detector for detecting motion intensity of the respective shots and a dynamic/static scene

detector for classifying the respective shots into a dynamic scene with much motions or a static scene with little motions based on the shot density and the motion intensity. Since the shot density accurately represents a quantity of motion of each scene, when the shot density is supposed to be a parameter, the scene can be accurately classified into the dynamic scene or the static scene.

(2) The apparatus is provided with an extractor for extracting a shot similar to a current target shot from shots after a shot before the target shot only by a predetermined interval; and a slow scene detector for classifying the target shot into a slow scene of the similar shot based on motion intensity of the target shot and the similar shot. Since the slow scene is similar to its reference scene and its motion intensity is lower than that of the reference scene, the slow scene can be reliably classified according to the above mentioned characteristics.

(3) A plurality of shots which continue near the slow scene or a scene with large audio signal are classified into a highlight scene. Since the highlight scene includes the slow scene or the scene with large audio signal, according to the above characteristic, the highlight scene can be classified out reliably.

(4) The apparatus is provided with detector for detecting a histogram relating to motion directions of the respective shots; and a detector for detecting a scene on which a camera

operation has been performed based on the histogram of motion direction. The histogram of motion direction shows characteristic distribution according to the camera operation, and in the case of a zooming scene, the histogram of motion direction becomes uniform, so that a number of elements of respective bins becomes larger than a reference number of elements. In the case of a panning scene, the histogram of the motion direction is concentrated in one direction (namely, only a number of elements of a certain bin becomes large), and the spatial distribution of motion becomes uniform. According to the above characteristic, therefore, the scene on which the camera operation has been performed can be classified reliably.

(5) The apparatus is provided with a detector for detecting shot density DS of the video; and a commercial scene detector for detecting a commercial scene based on the shot density. Since the shot density increases in the commercial scene, according to the above characteristics, the commercial scene can be classified reliably.

(6) The apparatus is provided with a detector for detecting a number of shot boundaries of the video; and a commercial scene detector for detecting a commercial scene based on the number of shot boundaries. Since the number of shot boundaries increases in the commercial scene, according to the above characteristics, the commercial scene can be classified reliably.

(7) When the video are compressed data, their motion

intensity, spatial distribution of motion and histogram of motion direction are detected by using a value of a motion vector of a predictive coding image existing in each shot. According to such a characteristic, types of the respective scenes of the compressed video can be detected on compressed domain without decoding.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram showing one embodiment of a scene classification apparatus.

FIG. 2 is a flowchart showing the scene classification procedure in the dynamic/static scene detector 1

FIG. 3 is a flowchart showing the scene classification procedure in the slow scene detector 2.

FIG. 4 is a flowchart showing the scene classification procedure in the first highlight scene detector 3.

FIG. 5 is a flowchart showing the procedure of the detecting process in the second highlight scene detector 4.

FIG. 6 is a flowchart showing the scene classification procedure in the highlight scene detector 5.

FIG. 7 is a pattern diagram showing the highlight scene classification method in the highlight scene detector 5.

FIG. 8 is a flowchart showing an operation of the video transition effect inserting section 6 .

FIG. 9 is a flowchart showing a procedure of the detecting process in the detector 7.

FIG. 10 is diagram showing a histogram distribution which is characteristic of a zooming scene.

FIG. 11 is diagram showing a histogram distribution which is characteristic of a zooming scene.

FIG. 12 is a flowchart showing a procedure of the detecting process in the panning scene detector 8.

FIG. 13 is diagram showing a histogram distribution which is characteristic of a panning scene.

FIG. 14 is a flowchart showing a procedure of the detecting process in the commercial scene detector 9.

FIG. 15 is a flowchart showing another procedure of the scene classification in the highlight scene detector 5.

Fig. 16 is a pattern diagram showing the highlight scene classification method, and hatched shots are the highlight scene.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Fig. 1 is a block diagram showing one embodiment of a scene classification apparatus to which the present invention is applied, and in this diagram, compressed or uncompressed video are classified into various scenes by shot unit.

A shot segmentation part 10 detects a cut from the video, and segments the video into shots based on the cut information. At the same time, audio information accompanied by the video or audio information multiplexed with the video is demultiplexed so that audio data are generated, and the audio data are distributed to an audio data analyzer 11.

The shot segmentation part 10 includes an NS detector 101 for detecting a total number of shots (NS) of the input video and a DS detector 102 for detecting shot density DS of the video

per a time unit.

A motion data analyzer 12 includes an IS detector 121 for detecting motion intensity (IS) of the video per unit region on the image, an MSD detector 122 for detecting spatial distribution of motion (MSD) of the video per unit region on the image, and an HD detector 123 for detecting a motion direction of the video per unit region on the image and generating a histogram of the obtained motion direction.

If the video are compressed data, the motion intensity IS, the spatial distribution of motion MSD and the histogram of motion direction HD can be detected by using a value of a motion vector of a predictive coding image existing in each shot. As to the motion intensity IS, "Motion Intensity" which is an element of "motion activity descriptor" defined in MPEG-7 or the like can be used.

On the other hand, if the video are uncompressed data, frames of the respective shots are compared by a block matching method or the like so that a change in the motion is predicted, and the change of the motion is expressed by a vector. Similarly to the above-mentioned manner, values of the motion intensity IS and the like are calculated by using the motion vector. At this time, as to the motion intensity IS as the shot, a value which is obtained by averaging a value of the motion intensity IS in the predictive coding image as an subject in a shot, a maximum value and an intermediate value of the values of the motion intensity IS can be used. As to the predictive coding

image to be a subject and the motion vector, any combination of a forward direction predictive coding image, a bidirectional predictive coding image, a forward directional motion vector and a backward directional motion vector in the bidirectional predictive coding image can be used.

As to the histogram of motion direction HD, directions of respective motion vectors are calculated in the motion information obtained as mentioned above, so that a histogram relating to all the motion vectors in the shot is structured. If a number of bins in the histogram is limited, it is desirable that the motion direction is quantized suitably.

This embodiment adopts quantization with a pitch of 45° to 8 directions which is treated in "Dominant Direction" as an element of the "motion activity descriptor" defined in MPEG-7, for example. Thresholding may be executed according to the size of the motion vector in such a manner that a motion vector which is not more than a certain level is not added as the element of the histogram.

The audio data analyzer 11 analyzes the audio data, and calculates an energy value E per audio signal or band. When the energy value E is calculated per band, an arbitrary band width can be selected, and weighting can be carried out per band.

A dynamic/static scene detector 1 classifies respective shots into a "dynamic" scene with much motions, a "static" scene with little motions and the other scene based on the shot density DS and the motion intensity IS.

A slow scene detector 2 extracts a shot similar to a current target shot from shots after a shot which is before the target shot by a predetermined interval. The slow scene detector 2 classifies the current shot into a slow scene of the similar shot based on the motion intensity of the current shot and the similar shot.

A first highlight scene detector 3 temporarily classifies a scene which is composed of a plurality of shots continuing just before the slow scene into the highlight scene. A second highlight scene detector 4 temporarily classifies a predetermined shot which continues before and after a highlight position detected based on the analyzed result of the audio data analyzer 11 into the highlight scene. A highlight scene detector 5 finally classifies the highlight scene based on temporary classified results of the respective highlight scenes. When a plurality of highlight scenes are connected, a video transition effect inserting section 6 inserts a video transition effect according to scene types between (before and after) the highlight scenes.

A zooming scene detector 7 classifies the respective shots into a scene on which zooming as one of camera operations has been performed or the other scenes based on the histogram of motion direction HD. A panning scene detector 8 classifies the respective shots into a scene on which panning as one of the other camera operations has been performed or the other scenes based on the histogram of motion direction HD and the spatial

distribution of motion MSD. A commercial scene detector 9 classifies the respective shots into a commercial scene or the other scenes based on the shot density DS.

The scene classification process in this embodiment will be detailed below according to a flowchart.

Fig. 2 is a flowchart showing the scene classification procedure in the dynamic/static scene detector 1, and this is executed on each shot segmented by the shot segmentation part 10. The shots are detected as the "dynamic" scene with much motions or the "static" scene with little motions.

At step S101, the shot density DS detected by the DS detector 102 of the shot segmentation part 10 is compared with first reference density DSref1, and the motion intensity IS detected by the IS detector 121 of the motion data analyzer 12 is compared with first reference intensity ISref1. When relationships: [DS > DSref1 and IS > ISref1] are satisfied, the sequence proceeds to step S102, so that the current target shot is classified into the "dynamic" scene. When [DS > DSref1 and IS > ISref1] are not satisfied at step S101, the sequence proceeds to step S103.

At step S103, the shot density DS is compared with second reference density DSref2 (< DSref1), and the motion intensity is compared with second reference intensity ISref2 (< Isref1). When [DS < DSref2 and IS < ISref2] are satisfied, the sequence proceeds to step S104, so that the current shot is classified into the "static" scene. When [DS < DSref2 and IS < ISref2] are not satisfied at step S104, the current shot is not classified

into the "dynamic" scene nor the "static" scene, so that this process is ended.

Fig. 3 is a flowchart showing the scene classification procedure in the slow scene detector 2, and decision is made whether the shots segmented by the shot segmentation part 10 are shots which compose a slow scene of another temporally previous shot.

At step S201, decision is made whether a shot S' which is similar to a current target shot S exists on an interval predated by predetermined time. When the shot S' similar to shots after the shot before the target shot S by the predetermined interval, the sequence proceeds to step S202.

In this embodiment, the slow scene detector 2 stores, for example, an image which is detected as a shot boundary by the shot segmentation part 10, namely, image data of a beginning image of the shot as a feature value of all the shots within a certain time, and stores image data of a reduced image of that image and a color layout descriptor which is obtained from the image and defined by MPEG-7, and makes detection whether the shot similar to the target shot input at current time exists in the past. The target shot is compared with not only the beginning image of the shot but also a middle of image of the shot and an image representing the shot (key frame).

At step S202, a differential value between the motion intensity IS of the target shot S and motion intensity IS' of the similar shot S' is compared with a reference differential

valueΔ ISref. When (IS' - IS) > Δ ISref is satisfied, the sequence proceeds to step S203, so that an interval (S' to S) between the target shot S and the similar shot S' is compared with a reference interval Δ Tref. When the shot S' is separated from the shot S by not less than the reference interval Δ Tref, the sequence proceeds to step S204 so that the target shot S is classified into the "slow" scene of the similar shot S'.

Fig. 4 is a flowchart showing the scene classification procedure in the first highlight scene detector 3, and the shot classified into the slow scene is used as a reference, so that a plurality of shots which continue just before the shot are provisionally classified into the highlight scene.

At step S301, decision is made whether a current target shot SO is classified into the "slow" scene. When the current target shot S0 is classified into the "slow" scene, the sequence proceeds to step S302, so that a predetermined number of shots S0-m to S0 (or for predetermined time), which continue just before the current shot S0, are combined so that the combined shots are classified into a "first highlight scene".

Fig. 5 is a flowchart showing the procedure of the detecting process in the second highlight scene detector 4, and when the audio signal or the energy value E obtained per band by the audio data analyzer 11 has a peak value, namely, the target shot S0 has stronger intensity than predetermined reference intensity, a plurality of shots which continue before and after the target shot S0 are provisionally classified into the highlight scene.

At step S401, decision is made whether an audio energy E accompanied by the current target shot S0 is a peak value based on a result of comparing with the energy value E of each shot input before the target shot S0. When the audio energy E is the peak value, the sequence proceeds to step S402, so that a difference between the energy value E of the target shot and the energy value E of the shot just before the target shot S0 is compared with the reference differential value Δ Eref. When the difference is larger than the reference differential value Δ Eref, the sequence proceeds to step S403, so that a predetermined number of shots S0-n to S0+n (or for predetermined time) which continue before and after the target shot S0 are combined and the combined shots are classified into a "second highlight scene".

Fig. 6 is a flowchart showing the scene classification procedure in the highlight scene detector 5, and the highlight scene is finally detected based on the detected results of the first and second highlight scene detectors 3, 4.

At step S501, decision is made whether a current target shot is classified into the first highlight scene, and when the current target shot is the first highlight scene, the sequence proceeds to step S502. At step S502, decision is made whether the current target shot is classified into the second highlight scene, and when it is the second highlight scene, the sequence proceeds to step S503. At step S503, the current target shot is classified into the highlight scene.

Fig. 7 is a pattern diagram showing the highlight scene classification method in the highlight scene detector 5, and hatched shots are the highlight scene. In this embodiment, only the shots which are shorted into both the first highlight scene and the second highlight scene are finally classified into the highlight scene.

Fig. 8 is a flowchart showing an operation of the video transition effect inserting section 6, and when a plurality of highlight scenes are extracted so as to be connected, the video transition effect is suitably selected so as to be inserted between the highlight scenes and played. When a video in which only the highlight scenes are collected is played, this process is executed in real time, or when a video in which only the highlight scenes are collected is created, this process is executed offline.

At step S601, decision is made whether the respective shots of the current target highlight scene are classified into the dynamic scene in the dynamic/static scene detector 1. When they are classified into the dynamic scene, a gradual transition such as instant image switching or wipe is inserted as a first video transition effect before the target highlight scene at step S602.

On the contrary, when the target highlight scene is classified into the static scene, the sequence proceeds from the step S603 to S604, so that an effect with large change in the image mixing ratio such as dissolve and fade is inserted as a second video transition effect. When the target highlight

scene is not classified into the dynamic scene nor the static scene, one of the first and the second video transition effects or a third video transition effect is inserted at step S605. When the shots composing the highlight scene include different scene types, one scene type is determined by majority.

Fig. 9 is a flowchart showing a procedure of the detecting process in the zooming scene detector 7, and Fig. 10 is diagram showing a histogram distribution which is characteristic of a zooming scene. In this process, decision is made whether the shots segmented by the shot segmentation part 10 are shots composing the zooming scene.

At step S701, as to a current target shot, dispersion of the histogram distribution DHD obtained by quantizing the histogram of motion direction HD is compared with a reference dispersion value DHDref. As shown in Fig. 10A, when the dispersion of the histogram distribution DHD is large and its value exceeds the reference dispersion value DHDref, the target shot is not classified into the zooming scene and this process is ended.

On the contrary, as shown in Fig. 10B, when the dispersion of the histogram distribution DHD is small and its value is smaller than the reference dispersion value DHDref, the sequence proceeds to step S702. At step S702, a number of elements of the bins in the histogram distribution N0 to N7 is compared with a reference number of elements Nref. As shown in Fig. 11A, when a number of the elements of any bins is less than the reference number

of elements Nref, the target shot is not classified into the zooming scene and the process is ended. As shown in Fig. 11B, when a number of the elements of all the bins N0 to N7 exceeds the reference number of elements Nref, the sequence proceeds to step S703, so that the current target shot is classified into the zooming scene.

Fig. 12 is a flowchart showing a procedure of the detecting process in the panning scene detector 8, and decision is made whether the shots segmented by the shot segmentation part 10 are shots composing a panning scene.

At step S801, as to a current target shot, decision is made whether the distribution of the histogram of motion direction HD detected by the HD detector 203 is concentrated in a certain specified bin (direction). As shown in Fig. 12, when only a number of elements of specified bin (here, a bin in a direction "4") is specifically large, the sequence proceeds to step S802, so that a number of the elements of the specified bin Nbinx is compared with the predetermined reference number of the elements Nref. As shown in Fig. 13, when a number of the elements of the specified bin Nbinx exceeds the reference number of the elements Nref, the sequence proceeds to step S803. At step S803, decision is made whether the spatial distribution of motion of the target shot MSD is uniform, and when it is uniform, the sequence proceeds to step S804 so that the current target shot is classified into the "panning" shot.

Fig. 14 is a flowchart showing a procedure of the detecting

process in the commercial scene detector 9, and when input video are television broadcasting program or the like including commercial, the shot density DS obtained by the shot segmentation part 10 is utilized so that decision is made whether a series of the input shots is a commercial scene.

At step S901, as to a current target shot, the shot density DS detected by the DS detector 102 is compared with reference density DSref. Not the shot density DS but a number of shot boundaries NC within a predetermined interval may be compared with a predetermined reference number NCref. When DS > DSref (NC > NCref), the sequence proceeds to step S902, so that the target shot is classified into the commercial scene.

Fig. 15 is a flowchart showing another procedure of the scene classification in the highlight scene detector 5, and the highlight scene is finally detected based on the detected results of the first and second highlight scene detectors 3, 4 and the detected result of the commercial scene detector 9.

In the highlight scene detecting method explained with reference to Figs. 4, 5, 6 and 7, the commercial scene having a comparatively high audio level is possibly misdecided as the highlight scene. In this embodiment, therefore, such misdecision is prevented and only the original highlight scene is classified reliably.

At step S511, decision is made whether a current target shot is classified into the first highlight scene, and when it is the first highlight scene, the sequence proceeds to step S512.

At step S512, decision is made whether the target shot is classified into the second highlight scene, and when it is the second highlight scene, the sequence proceeds to step S513. At step S513, decision is made whether the current target shot is the commercial scene, and when it is not the commercial scene, the sequence proceeds to step S514 so that the current target shot is classified into the highlight scene.

Fig. 16 is a pattern diagram showing the highlight scene classification method, and hatched shots are the highlight scene. In this embodiment, only the shots other than the first highlight scene, the second highlight scene and the commercial scene are finally classified into the highlight scene.

According to the present invention, the scenes of the uncompressed or compressed video are classified into various types, so that a desired scene can be searched and viewed from the video and a lot of video can be effectively classified, and also the classified scene can be played in an optimum form.